CHAPTER 25

SUPPLY AND THE COSTS OF PRODUCTION

Product prices are determined by the interaction of the forces of demand and supply. Preceding chapters have focused our attention upon the factors underlying demand. The basic factor underlying the ability and willingness of firms to supply a product in the market is the cost of production. The production of any good requires the use of economic resources which, because of their relative scarcity, bear price tags. The amount of any product which a firm is willing to supply in the market depends upon the prices, or costs, of the resources essential to its production, on the one hand, and the price which the product will bring in the market, on the other. The present chapter is concerned with the general nature of production costs. Product prices are introduced in the following several chapters, and the supply decisions of producers are then explained.

ECONOMIC COSTS IN REAL AND MONEY TERMS

Economic costs are those payments which must be received by resource owners in order to assure that they will continue to supply them in a particular line of production. This definition goes back to the basic fact that resources are scarce and have alternative uses. First in real (physical) terms, the economist's definition of costs simply suggests that to use a resource in producing one product entails giving up some alternative product. The real cost of producing 5 units of X is the number of units of Y or Z which the resources used in X could otherwise have produced. The production possibilities curve of Chapter 2 clearly embodies this notion of costs. Note at point *C* in Table 2–1 that the real cost of producing 100,000 more units of bread is the 3,000-unit decrease in the production of drill presses which will necessarily be entailed. A final illustration: Suppose an assembly-line worker can be used in producing automobiles, washing machines, and refrigerators. The real cost of using this worker in producing automobiles is the contribution which he might otherwise have made in the production of washing machines and refrigerators.

When expressed in money terms, the notion of economic costs is a bit more elusive. The main reason for this is that we, as the accountant or businessman, typically think of costs as being essentially money payments, that is, cash outlays, which a firm makes to the "outsiders" who supply labor services, materials, fuel, transportation services, power, and so forth, to the firm. These expenditures, or *explicit costs,* are certainly a part of the economist's definition of costs; but they are only a part. The economist would also include any nonexpenditure, or *implicit,* costs, that is, the value of any resources which are owned and employed by an enterprise. The economist's reasoning is simple and very pertinent: Regardless of whether a resource is owned or hired by an enterprise, there is a cost involved in using that resource in a specific employment. In real terms that cost is the units of alternative products which are forgone. In money terms it is the money payment which the self-employed resource could have earned in its best alternative employments. For example, suppose Brooks operates a corner grocery as a sole proprietor. He owns his store building and supplies all his own labor and money capital. Though his enterprise has no explicit rental or wage costs, implicit rents and wages are incurred. By using his own building for a grocery, Brooks sacrifices the $200 monthly rental income which he could otherwise have earned by renting it to someone else. Similarly, by using his money capital and labor in his own enterprise, Brooks sacrifices the interest and wage incomes which he otherwise could have earned by supplying these resources in their best alternative employments. And, finally, by running his own enterprise, Brooks forgoes the earnings he could realize by supplying his entrepreneurial efforts in someone else's firm.

The minimum payment required to keep Brooks's entrepreneurial talents engaged in this enterprise is sometimes called a *normal profit.* As implicit rent or implicit wages, this normal return for the performing of entrepreneurial functions is an implicit cost. If this minimum, or normal, return is not realized, the entrepreneur will withdraw his efforts from this line of production and reallocate them to some alternative line of production. Or the individual may cease being an entrepreneur in favor of becoming a laborer.

In short, the economist includes as costs all payments—explicit and implicit, the latter including a normal profit—required to retain resources in a given line of production.

Economic, or Pure, Profits

Our discussion of economic costs correctly suggests that economists and accountants use the term "profits" differently. By "profits" the accountant generally means total receipts less explicit costs. But to the economist "profits" means total receipts less *all* costs (explicit and implicit, the latter including a normal return to the entrepreneur). Therefore, when an economist says that a firm is just covering its costs, he means that all explicit and implicit costs are being met and that the entrepreneur is therefore receiving a return just large enough to retain his talents in his present line of production. If a firm's total receipts exceed all its economic costs, any residual accrues to the entrepreneur. This residual is called an *economic,* or *pure, profit.* It is not a cost, because by definition it is a return in excess of the normal profit required to retain the entrepreneur in this particular line of production. In Chapter 32 we shall find that economic profits are associated with risk bearing and monopoly power.

Short Run and Long Run

The costs which a firm or industry incurs in producing any given output will depend upon the types of adjustment it is able to make in the amounts of the various resources it employs. The quantities employed of many resources—labor, raw materials, fuel, power, and so forth—can be varied easily and quickly. But the amounts of other resources demand more time for adjustment. For example, the capacity of a manufacturing plant, that is, the size of the factory building and the amount of machinery and equipment therein, can only be varied over a considerable period of time. In some heavy industries it may take several years to alter plant capacity.

These differences in the time necessary to vary the quantities of the various resources used in the productive process make it essential to distinguish between the short run and the long run. The *short run* refers to a period of time, too short to permit an enterprise to alter its plant capacity yet long enough to permit a change in the level at which the fixed plant is utilized. The firm's plant capacity is fixed in the short run, but output can be varied by applying larger or smaller amounts of manpower, materials, and so forth, to that plant. Existing plant capacity can be used more or less intensively in the short run.

From the viewpoint of existing firms the *long run* refers to a period of time long enough to allow these firms to change the quantities of *all* resources employed, including plant capacity. From the viewpoint of an industry the long run also encompasses enough time for existing firms to dissolve and leave the industry and for new firms to be created and enter the industry. While the short run is a "fixed-plant" time period, the long run is a "variable-plant" time period.

Some examples will make clear the distinction between the short run and the long run. If a General Motors plant were to hire an extra 100 workers or to add an entire shift of workers, these would be short-run adjustments. If the same GM plant were to add a new wing to its building and install more equipment, this would be a long-run adjustment. Studebaker's abandonment of its South Bend, Indiana, plant in 1964 was a long-run adjustment.

It is important to note that the short run and the long run are conceptual rather than specific calendar time periods. In light manufacturing industries, changes in plant capacity may be negotiated almost overnight. A small firm making men's clothing can increase its plant capacity in a few days or less simply by ordering and installing a couple of new cutting tables and several extra sewing machines. But heavy industry is a different story. It may take Ford or General Motors several years to construct a new assembly plant and to install elaborate assembly-line equipment.

We turn now to the task of analyzing production costs in the short-run, or fixed-plant, period. Following this we consider costs in the long-run, or variable-plant, period.

PRODUCTION COSTS IN THE SHORT RUN

A firm's costs of producing any output will depend not only upon the prices of needed resources, but also upon technology—the quantity of resources it takes to produce that output. It is the latter, technological aspect of costs with which we are concerned for the moment. In the short run a firm can change its output by adding variable resources to a fixed plant. Question: How does output change as more and more variable resources are added to the firm's fixed resources?

Law of Diminishing Returns

The answer is provided in general terms by the *law of diminishing returns*. This engineering law states that *as successive units of a variable resource (say, labor) are added to a fixed resource (capital), beyond some point the extra, or marginal, product attributable to each additional unit of the variable resource will decline*. Stated somewhat differently, if additional workers are applied to a given amount of capital equipment, as is the case in the short run, eventually output will rise less than in proportion to the increase in the number of workers employed. A couple of examples will illustrate this law.

Suppose a farmer has a fixed amount of land—say, 80 acres—which he has planted in corn. Assuming the farmer does not cultivate his cornfields at all, his yield will be, say, 40 bushels per acre. If he cultivates the land once, output may rise to 50 bushels per acre. A second cultivation may increase output to 57 bushels per acre, a third to 61, and a fourth to, say, 63. But further cultivations will add little or nothing to total output. Successive cultivations add less and less to the land's yield. If this were not the case, the world's needs for corn could be fulfilled by extremely intense cultivation of this single 80-acre plot of land. Indeed, if diminishing returns did not occur, the world could be fed out of a flowerpot.

• The law of diminishing returns also holds true in nonagricultural industries. Assume a small planing mill is manufacturing unupholstered furniture. The mill has a given amount of equipment in the form of lathes, planers, saws, sanders, and so forth. If this firm hired just one or two workers, total output and production per man would be very low. These workers would have a number of different jobs to perform, and the advantages of specialization would be lost. Time would also be lost in switching from one job operation to another, and the machines would stand idle most of the time. In short, the plant would be undermanned, and production therefore would be inefficient. These difficulties would disappear as more workers were added. Equipment would be more fully utilized, and workers could now specialize on a single job. Thus as more workers are added to the initially undermanned plant, the extra or marginal product of each will tend to rise as a result of more efficient production. But this cannot go on indefinitely. As still more workers are added, problems of overcrowding will arise. Workers must wait in line to use the machinery, so now *workers* are underutilized. The extra, or marginal, product of additional workers declines because the plant is overmanned. In the extreme the continuous addition of labor to the plant would use up all standing room, and production would be brought to a standstill!

Table 25–1 illustrates the law of diminishing returns numerically. In this instance diminishing marginal product is incurred with the hire of the third worker. Total product is found by simply accumulating the extra, or marginal, product attributable to each successive worker. Total product will increase so long as marginal product is positive.

Fixed, Variable, and Total Costs

The production data described by the law of diminishing returns must be coupled with resource prices to determine the total and per unit costs of producing various outputs. We have already emphasized that in the short run some resources—those associated with the firm's plant—are fixed. Others are variable. This correctly suggests that in the short run costs can be classified as either fixed or variable.

Fixed costs are those costs which in total do not vary with changes in output. Fixed costs are associated with the very existence of a firm's plant and therefore must be paid even if the firm's rate of output is zero. Such costs as interest on a firm's bonded indebtedness, rental payments, a portion of depreciation on equipment and buildings, insurance premiums, and the salaries of top management and key personnel are generally fixed costs. In column 2 of Table 25–2

TABLE 25–1. THE LAW OF DIMINISHING RETURNS (hypothetical data)

(1) Inputs of the variable resource (labor)	(2) Extra, or marginal, product	(3) Total product
0		0
	5	
1		5
	8	
2		13
	5	
3		18
	4	
4		22
	3	
5		25
	2	
6		27

TABLE 25–2. TOTAL- AND AVERAGE-COST SCHEDULES FOR AN INDIVIDUAL FIRM IN THE SHORT RUN (hypothetical data)

Total-cost data, per week				Average-cost data, per week			
(1) Total product	(2) Total fixed cost	(3) Total variable cost	(4) Total cost	(5) Average fixed cost, or (2) ÷ (1)	(6) Average variable cost, or (3) ÷ (1)	(7) Average total cost, or (4) ÷ (1)	(8) Marginal cost, or Δ(4)
0	$100	$ 0	$ 100				
1	100	90	190	$100.00	$90.00	$190.00	$ 90
2	100	170	270	50.00	85.00	135.00	80
3	100	240	340	33.33	80.00	113.33	70
4	100	300	400	25.00	75.00	100.00	60
5	100	370	470	20.00	74.00	94.00	70
6	100	450	550	16.67	75.00	91.67	80
7	100	540	640	14.29	77.14	91.43	90
8	100	650	750	12.50	81.23	93.73	110
9	100	780	880	11.11	86.67	97.78	130
10	**100**	**930**	**1,030**	**10.00**	**93.00**	**103.00**	150

we have assumed that the firm's fixed costs are $100. Note that this fixed-cost figure prevails at all levels of output, including zero.

Variable costs are those costs which increase with the level of output. Variable costs include payments for labor, materials, fuel, power, transportation services, and similar variable resources. In column 3 of Table 25–2 we find that the total of variable costs changes with output, but note that the rate of increase in variable costs is not constant. As production begins, variable costs will for a time increase at a decreasing rate; this is true through the fourth unit of output. Beyond the fourth unit, however, variable costs increase at an increasing rate. The explanation of this behavior of variable costs lies in the law of diminishing returns. Because of increasing marginal product, smaller and smaller increases in the amounts of variable resources will be needed for a time to get successive units of output produced. This means that total variable costs will increase at a decreasing rate. But when marginal product begins to decline as diminishing returns are encountered, it will be necessary to use larger and larger additional amounts of variable resources to produce each successive unit of output. Total variable costs will therefore increase at an increasing rate.

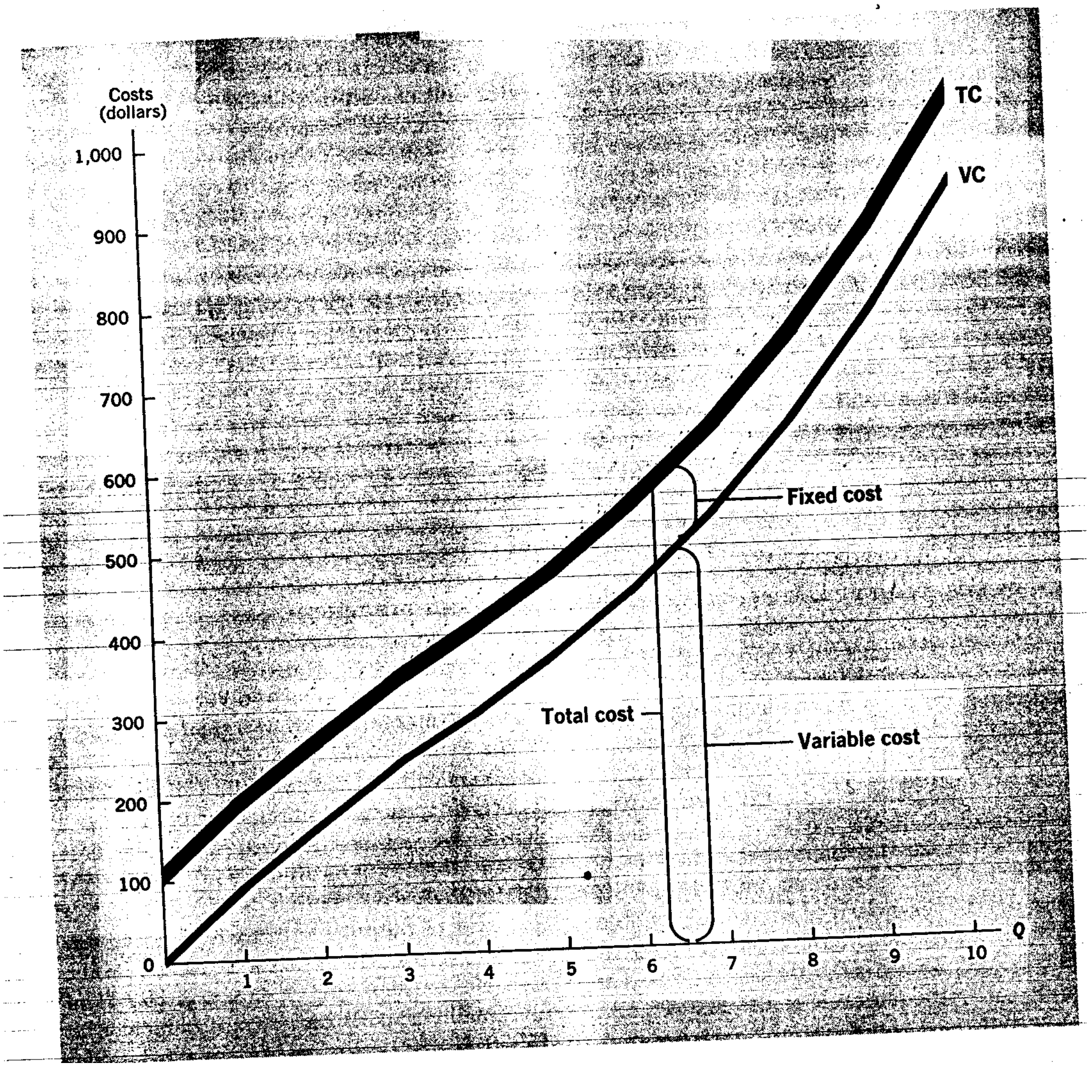
Total cost is self-defining: it is the *sum of fixed and variable costs at each level of output.* It is shown in column 4 of Table 25–2.

At zero units of output total cost is equal to the firm's fixed costs. Then for each unit of production—1 through 10—total cost varies at the same rate as does variable cost.

Figure 25–1 shows graphically the fixed-, variable-, and total-cost data of Table 25–2. Note that total variable cost is measured from the horizontal axis and total fixed cost is added vertically to total variable cost in locating the total cost curve.

The distinction between fixed and variable costs is of no little significance to the businessman. Variable costs are those costs which the businessman can control or alter in the short run by changing his level of production. On the other hand, fixed costs are

FIGURE 25–1. TOTAL COST IS THE SUM OF FIXED AND VARIABLE COSTS. Variable costs (VC) vary with output. Fixed costs are independent of the level of output. The total cost (TC) of any output is the (vertical) sum of the fixed and variable costs of that output.



clearly beyond the businessman's control; such costs are incurred and must be paid regardless of output level.

Per Unit, or Average, Costs

Producers are certainly interested in their total costs, but they are equally concerned with their per unit, or average, costs. In particular, average-cost data are more usable for making comparisons with product price, which is always stated on a per unit basis. Average fixed cost, average variable cost, and average total cost are shown in columns 5 to 7 of Table 25–2. It is important that we know how these unit-cost figures are derived and how they vary as output changes.

Average fixed cost (AFC) is found by dividing total fixed costs by the corresponding output. AFC declines as output increases. Whereas total fixed costs are, by definition, independent of output, AFC will decline as output increases. As output increases, a given total fixed cost of $100 is obviously being spread over a larger and larger output. When output is just 1 unit, total fixed costs and AFC are equal—$100. But at 2 units of output, total fixed costs of $100 become $50 worth of fixed costs per unit; then $33.33, as $100 is spread over 3 units; $25, when spread over 4 units; and so forth. This is what businessmen commonly refer to as "spreading the overhead." We find in Figure 25–2 that AFC graphs as a continually declining figure as total output is increased.

Average variable cost (AVC) is found by dividing total variable cost by the corresponding output. AVC declines initially, reaches a minimum, and then increases again. Graphically, this provides us with a U-shaped AVC curve, as is shown in Figure 25–2.

Because total variable cost reflects the law of diminishing returns, so must the AVC figures, which are derived from total variable cost. Because of increasing returns it takes fewer and fewer additional variable resources to produce each of the first 4 units of output. As a result, variable cost per unit will decline. AVC hits a minimum with the fifth unit of output, and beyond this point AVC rises as diminishing returns necessitate the use of more and more variable resources to produce each additional unit of output. In more direct terms, at low levels of output production will be relatively inefficient and costly, because the firm's fixed plant is undermanned. Not enough variable resources are being combined with the firm's plant; production is inefficient, and per unit variable costs are therefore relatively high. As output expands, however, greater specialization and a more complete utilization of the firm's capital equipment will make for more efficient production. As a result, variable cost per unit of output will decline. As more and more variable resources are added, some point will eventually be reached where diminishing returns are incurred. The firm's capital equipment will now be overmanned, and the resulting overcrowding and overutilization of machinery impairs efficiency. This means that AVC will increase.

Average total cost (ATC) can be found by dividing total cost by total output or, more simply, by adding AFC and AVC for each of the ten levels of output. These data are shown in column 7 of Table 25–2. Graphically, ATC is found by adding vertically the AFC and AVC curves, as in Figure 25–2. Thus the vertical distance between the ATC and AVC curves reflects AFC at any output.

Marginal Cost

There remains one final and very crucial cost concept—marginal cost. *Marginal cost (MC) is the extra, or additional, cost of producing one more unit of output.* MC can be determined for each additional unit of output simply by noting the change in total cost which that unit's production entails. In Table 25–2 we find that production of the first unit of output increases total cost from $100 to $190. Therefore, the additional, or marginal, cost of that first unit is $90. The marginal cost of the second unit is $80 ($270 − $190); the MC of the third is $70

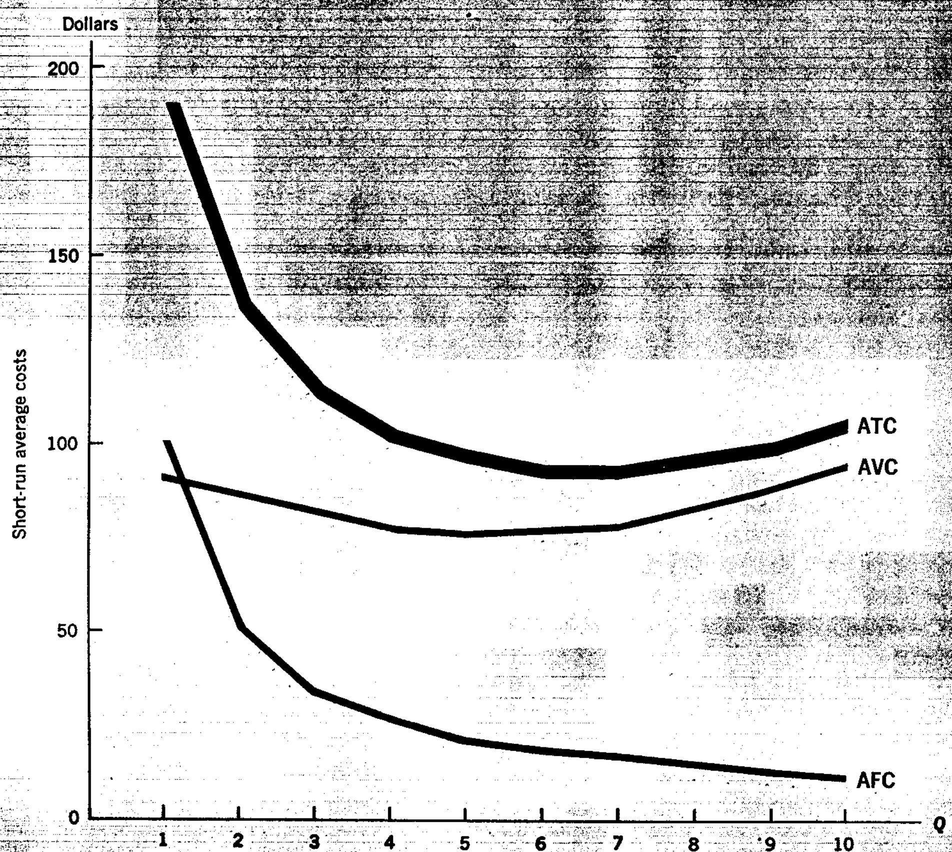


FIGURE 25–2. THE AVERAGE-COST CURVES.
Average total cost (ATC) is the vertical sum of average variable cost (AVC) and average fixed cost (AFC). AFC necessarily falls as a given amount of fixed costs is apportioned over a larger and larger output. AVC initially falls because of increasing physical returns but then rises because of diminishing physical returns.

($340 – $270); and so forth. MC for each of the ten units of output is shown in column 8 of Table 25–2. MC can also be calculated from the total-variable-cost column. Why? Because the only difference between total cost and total variable cost is the constant amount of fixed costs. Hence, the change in total cost and change in total variable cost associated with each additional unit of output is the same.

Marginal cost is a very strategic concept, because it designates those costs over which the firm has the most direct control. More specifically, MC indicates those costs which are incurred in the production of the last unit of output and, simultaneously, the cost which can be "saved" by reducing total output by the last unit. Average-cost figures do not provide this information. For example, suppose the firm is undecided as to whether it should produce 3 or 4 units of output. At 4 units of output Table 25–2 indicates that ATC is $100. But the firm does not increase its total costs by $100 by producing, nor does it "save" $100 by not producing, the fourth unit. Rather the change in costs involved here is only $60, as the MC column of Table 25–2 clearly reveals. A firm's decisions as to what output to produce are marginal decisions, that is, decisions to produce a few more or a few less units. Marginal cost reveals the change in costs which one more unit or one less unit of output entails. When coupled with marginal revenue, which we found in Chapter 23 indicates the change in revenue from one more or one less unit of output, marginal cost allows a firm to determine whether it is profitable to expand or contract its level of production. The analysis in the next four chapters centers upon these marginal calculations.

Marginal cost is shown graphically in Figure 25–3. Note that marginal cost declines sharply, reaches a minimum, and then rises rather sharply. This mirrors the fact that variable cost, and therefore total cost, increases first at a decreasing rate and then at an increasing rate (see Figure 25–1 and columns 3 and 4 of Table 25–2). This, you will recall, is in accord with the law of diminishing returns.

Furthermore, it is notable that marginal cost cuts both AVC and ATC at their minimum points. This marginal-average relationship is a matter of mathematical necessity, which a common-sense illustration can make readily apparent. Suppose a baseball pitcher has allowed his opponents an average of 3 runs per game in the first three games he has pitched. Now whether his average falls or rises as a result of pitching a fourth (marginal) game will depend upon whether the additional runs he allows in that extra game are fewer or more than his current 3-run average. If he allows fewer than 3 runs—for example, 1—in the fourth game, his total runs will rise from 9 to 10, and his average will fall from 3 to 2½ (10 ÷ 4). Conversely, if he allows more than 3 runs—say, 7—in the fourth game, his total will rise from 9 to 16 and his average from 3 to 4 (16 ÷ 4). So it is with costs. When the amount added to total cost (marginal cost) is less than the average of total cost, ATC will fall. Conversely, when marginal cost exceeds ATC, ATC will rise. This means in Figure 25–3 that so long as MC lies below ATC, the latter will fall, and where MC is above ATC, ATC will rise. Therefore at the point of intersection where MC equals ATC, ATC has just ceased to fall but has not yet begun to rise. This, by definition, is the minimum point on the ATC curve. Because MC can be defined as the addition either to total cost or to total variable cost resulting from one more unit of output, this same rationale explains why MC also cuts AVC at the latter's minimum point. No such relationship exists for MC and average fixed cost, because the two are simply not related; marginal cost embodies only those costs which change with output, and fixed costs by definition are independent of output.

PRODUCTION COSTS IN THE LONG RUN

In the long run all desired resource adjustments can be negotiated by an industry and the individual firms which it comprises. The firm can alter its plant capacity; it can build a larger plant or revert to a smaller plant than that assumed in Table 25–2. The industry can also change its plant size; the long run is an amount of time sufficient for new firms to enter or old firms to leave an industry. The impact of the entry and exodus of firms from an industry will be discussed in the next chapter; here we are concerned only with changes in plant capacity made by a single firm. And in considering these adjustents, we couch our analysis in terms of ATC, making no distinction between fixed and variable costs for the obvious reason that all resources and therefore all costs are variable in the long run.

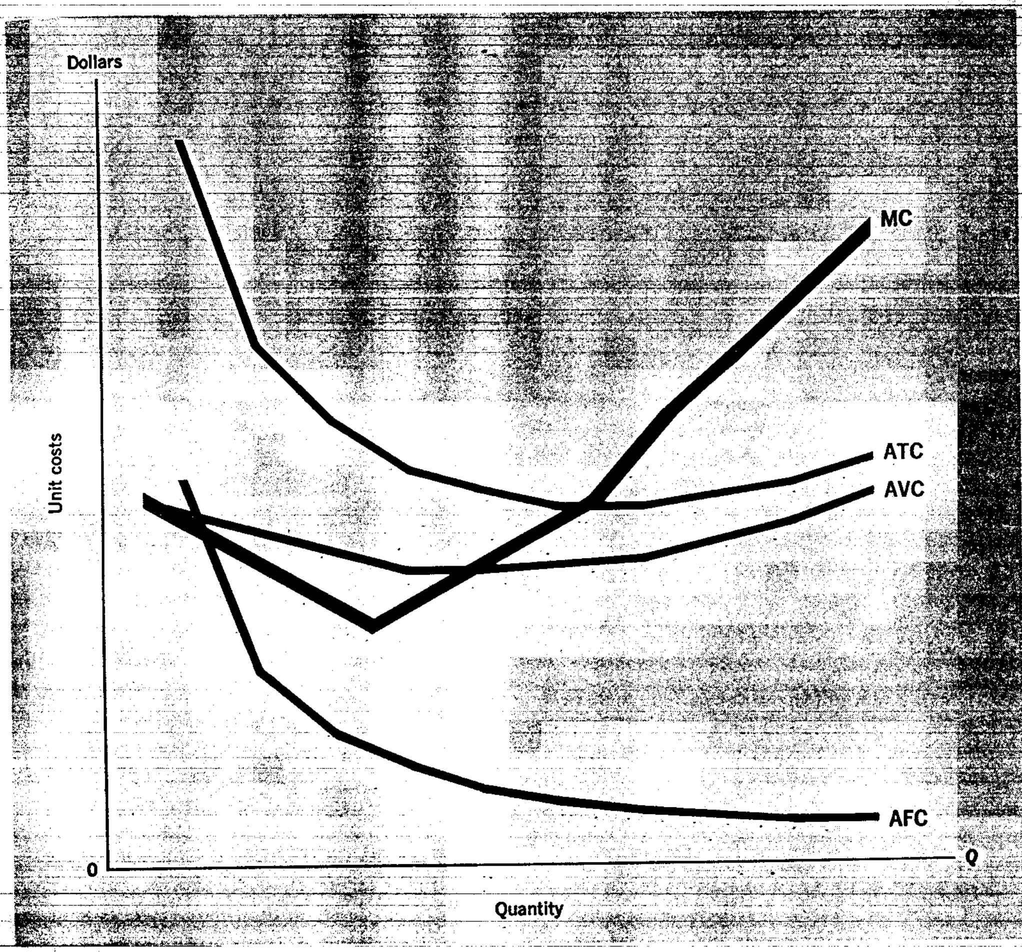


FIGURE 25–3. THE RELATIONSHIP OF MARGINAL COST TO AVERAGE TOTAL COST AND AVERAGE VARIABLE COST.
Marginal cost (MC) cuts both ATC and AVC at their minimum points. This is so because whenever the extra or marginal amount added to total cost (or variable cost) is less than the average of that cost, the average will necessarily fall. Conversely, whenever the marginal amount added to total (or variable) cost is greater than the average of total cost, the average must rise.

Suppose a single-plant manufacturing enterprise starts out on a small scale and then, as the result of successful operations, expands to successively larger plant sizes. What will happen to average total costs as this growth occurs? The answer is this: For a time successively larger plants will bring lower average total costs. However, eventually the building of a still larger plant will cause ATC to rise.

Figure 25–4 illustrates this situation for five possible plant sizes. ATC-1 is the average-total-cost curve for the smallest of the five plants, and ATC-5 for the largest. The relationship of the five plant sizes to one another is clearly that stated above. Con-

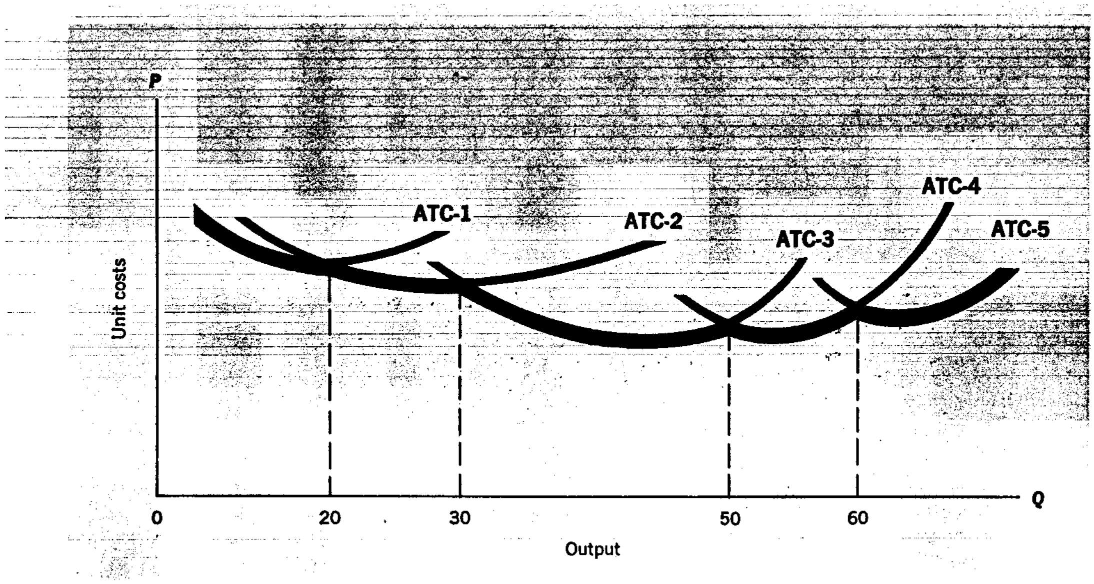


FIGURE 25–4. THE LONG-RUN AVERAGE-COST CURVE: FIVE POSSIBLE PLANT SIZES.
The long-run average-cost curve is made up of segments of the short-run cost curves (ATC-1, ATC-2, etc.) of the various-sized plants from which the firm might choose. Each point on the bumpy planning curve shows the least unit cost attainable for any output when the firm has had time to make all desired changes in its plant size.

structing a larger plant will entail lower per unit costs through plant size 3. But beyond this point a larger plant will mean a higher level of average total costs.

The dotted lines perpendicular to the output axis are crucial. They indicate those points at which the firm should change plant size in order to realize the lowest attainable per unit costs of production. To illustrate in terms of Figure 25–4: For all outputs up to 20 units the lowest per unit costs are attainable with plant size 1. However, if the firm's volume of sales expands to some level greater than 20 but less than 30 units, it can achieve lower per unit costs by constructing a larger plant—plant size 2. For any output between 30 and 50 units plant size 3 will yield the lowest per unit costs. For the 50–60-unit range of output, plant size 4 must be built to achieve the lowest unit costs. Lowest per unit costs for any output in excess of 60 units demand the construction of the still larger plant of size 5.

Tracing these adjustments, we can conclude that the long-run ATC curve for the enterprise will comprise segments of the short-run ATC curves for the various plant sizes which can be constructed. *The long-run ATC curve shows the least per unit cost at which any output can be produced after the firm has had time to make all appropriate adjustments in its plant size.* In Figure 25–4 the heavy, bumpy curve is the firm's long-run ATC curve or, as it is often called, the firm's planning curve. In most lines of production the choice of plant sizes is much wider than that assumed in our illustration. In fact, in many industries the number of possible plant sizes is virtually unlimited. This means that in time very small changes in the volume of output (sales) will prompt appropriate changes in the size of the plant.

Graphically this means that the planning curve will be smooth rather than bumpy. Figure 25–5 is illustrative.

Economies and Diseconomies of Scale

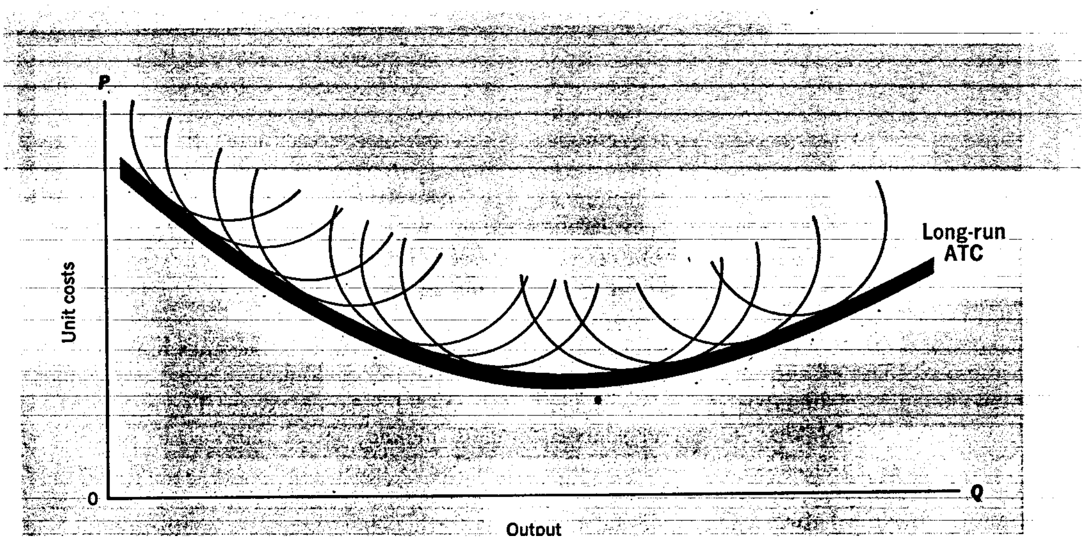
We have patiently accepted the contention that for a time a larger and larger plant size will entail lower unit costs but that beyond some point successively larger plants will mean higher average total costs. Now we must explain this point. Exactly why is the long-run ATC curve U-shaped? It must be emphasized, first of all, that the law of diminishing returns is not applicable here, because it presumes that one resource is fixed in supply and, as we have seen, the long run assumes that all resources are variable. What then is our explanation? The U-shaped long-run average-cost curve is explainable in terms of what economists call "economies and diseconomies" of large-scale production.

Economies of large scale. Economies of scale or, more commonly, economies of mass production, explain the downsloping part of the long-run ATC curve. As the size of a plant increases, a number of considerations will for a time give rise to lower average costs of production.

1. Increased specialization in the use of labor is feasible as a plant increases in size. The hire of more workers means that jobs can be divided and subdivided. Instead of performing five or six distinct operations in the productive process, each worker may now have just one task to perform. Workers can be used full time on those particular operations at which they have special skills. In a small plant a skilled machinist may spend half his time performing unskilled tasks. This makes for high production costs. Further, the dividing of work operations which large scale allows will give workers the opportunity to become very proficient at

FIGURE 25–5. THE LONG-RUN AVERAGE-COST CURVE: UNLIMITED NUMBER OF PLANT SIZES.

If the number of possible plant sizes is very large, the long-run average-cost curve approximates a smooth curve. Economies and diseconomies of scale cause the curve to be U-shaped.



the specific tasks assigned them. The Jack-of-all-trades who is burdened with five or six jobs will not be likely to become very efficient in any of them. When allowed to concentrate on one task, the same worker may become highly efficient. Finally, greater specialization tends to eliminate the loss of time which accompanies the shifting of workers from one job to another.

2. *Large-scale production also permits better utilization of, and greater specialization in, management.* A foreman capable of handling fifteen or twenty men will be underutilized in a small plant hiring only eight or ten men. The production staff can be doubled with no increase in administrative costs. In addition, small firms will not be able to use management specialists to best advantage. In a small plant a sales specialist may be forced to divide his time between several executive functions—for example, sales, personnel, and finance. A larger scale of operations will mean that the sales expert can devote full time to supervising sales while appropriate specialists are added to perform other managerial functions. Greater efficiency and lower unit costs are the net result.

3. *Small firms are often not able to utilize the most efficient productive equipment.* In many lines of production the most efficient machinery is available only in very large and extremely expensive units. Furthermore, effective utilization of this equipment demands a high volume of production. This means only large-scale producers are able to afford and operate efficiently the best available equipment.

To illustrate: In the automobile industry the most efficient fabrication method entails the use of extremely elaborate assembly-line equipment. The efficient use of this equipment demands an annual output of thousands of automobiles per year. Only very large-scale producers can afford to purchase and use this equipment efficiently. The small-scale producer is between the devil and the deep blue sea. To fabricate automobiles with the use of other equipment is inefficient and therefore more costly per unit. The alternative of purchasing the most efficient equipment and underutilizing it with a small level of output is equally inefficient and costly.

4. *The large-scale producer is in a better position to utilize by-products than is a small firm.* The large meat-packing plant makes glue, fertilizer, pharmaceuticals, and a host of other products from animal remnants which would be discarded by smaller producers.

All these technological considerations—greater specialization in the use of labor and management, the ability to use the most efficient equipment, and the effective utilization of by-products—will contribute to lower unit costs for the smaller producer who is able to expand his scale of operations.

Diseconomies of large scale. But in time the expansion of a firm will likely give rise to diseconomies and therefore higher per unit costs.

The main factor causing diseconomies of scale has to do with certain managerial problems which typically arise as a firm becomes a large-scale producer. In a small plant a single key executive may render all the basic decisions relative to his plant's operation. Because of the firm's smallness he is close to the production line. He can therefore comprehend the various aspects of the firm's operations and digest the information fed to him by his subordinates to the end that efficient decision making is possible.

This neat picture changes, however, as a firm grows. The management echelons between the executive suite and the assembly line become many; top management is far removed from the actual production operations of the plant. It becomes impossible for one man to assemble, understand, and digest all the information essential to rational decision making in a large-scale enterprise. Authority must be delegated to innumerable vice-presidents, second vice-presidents, and so forth. This expansion in the depth and width of management entails problems of coordination and bureaucratic red tape which can eventually impair the efficiency of a firm and lead to higher costs.

Significance of economies and diseconomies of scale. Economies and diseconomies of scale are something more than a plausible pipedream of economic theorists. Indeed, in most American manufacturing industries economies of scale have been of great significance. Firms which have been able to expand their scale of operations to realize the economies of mass production have survived and flourished. Those unable to achieve this expansion have found themselves in the unenviable position of high-cost producers, doomed to a marginal existence or ultimate insolvency.

Diseconomies of scale, when encountered, can be equally significant. The organizational structure of General Motors, for example, is designed to avoid managerial diseconomies which its gigantic size would otherwise entail. This industrial colossus has subdivided itself into some thirty-four operating subdivisions, each of which is basically autonomous and in some cases—for example, its five automobile-producing divisions (Chevrolet, Buick, Oldsmobile, Pontiac, and Cadillac)—competing. A degree of decentralization has been sought which will allow full realization of the economies of mass production yet help to avoid diseconomies of scale.[1] Another example: Some economists feel that U.S. Steel has declined in relative importance in the steel industry because of diseconomies of scale. One authority has described U.S. Steel as[2]

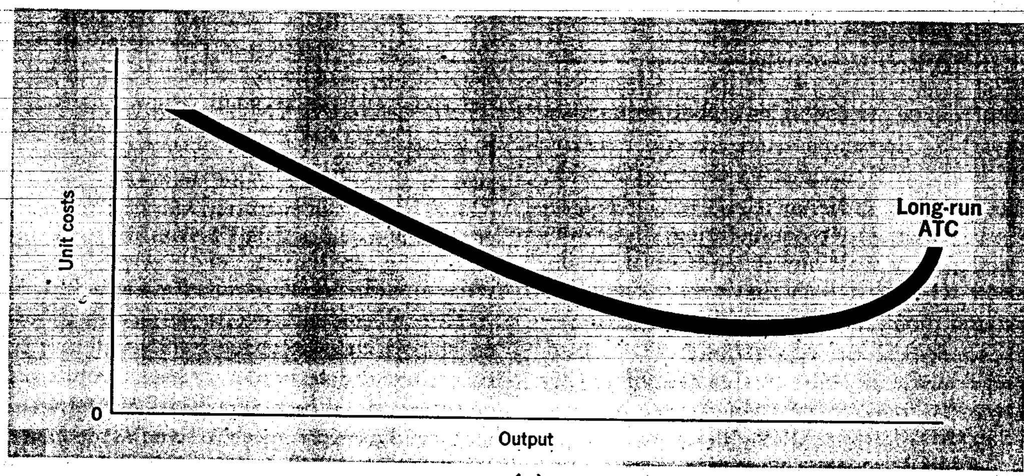
> . . . a big sprawling inert giant, whose production operations were improperly coordinated; suffering from a lack of a long-run planning agency; relying on an antiquated system of cost accounting: with an inadequate knowledge of the costs or of the relative profitability of the many thousands of items it sold; with production and cost standards generally below those considered everyday practice in other industries; with inadequate knowledge of its domestic markets and no clear appreciation of its opportunities in foreign markets; with less efficient production facilities than its rivals had; slow in introducing new processes and new products.

These comments correctly imply that economies and diseconomies of scale are a fundamental determinant of the structure of any industry. Where economies of scale are many and diseconomies are remote, the long-run ATC curve will decline over a long range of output as in Figure 25–6*a*. Such is the case in the automobile, aluminum, steel, and a host of other heavy industries. This means that, given consumer demand, efficient production will be achieved only with a small number of large producers. On the other hand, where economies of scale are few and diseconomies quickly encountered, minimum unit costs will be achieved at a modest level of production. The long-run ATC curve for such a situation is shown in Figure 25–6*b*. In such industries a given level of consumer demand will support a large number of relatively small producers. Many of the retail trades and some types of farming fall into this category. So do certain types of light manufacturing, for example, the baking, clothing, and shoe industries. Fairly small firms are as efficient as, or more efficient than, large-scale producers in such industries.

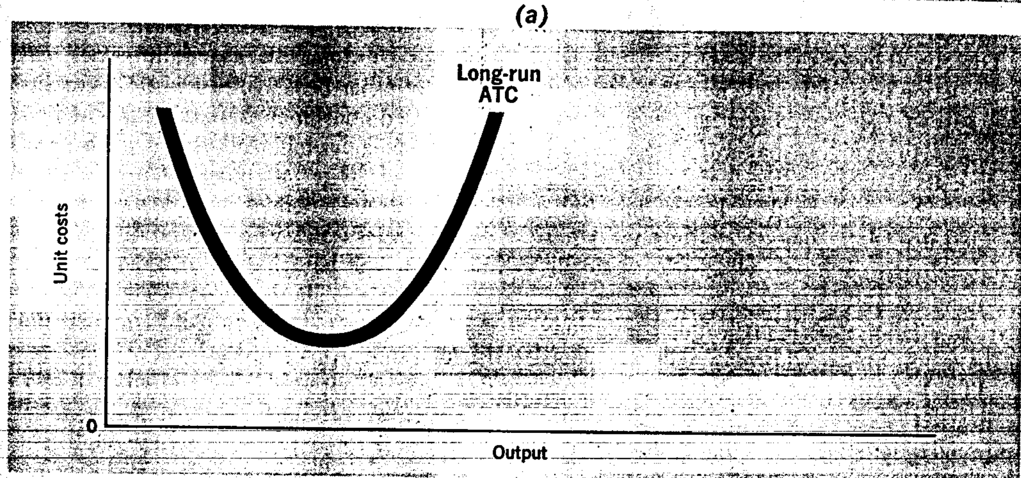
In some industries we find a mixture of large and small producers operating with roughly the same degree of efficiency—the meat-packing, household-appliance, and furniture industries are representative. In such industries the long-run ATC curve may be such that there exists a wide range of output between the point at which available economies of scale are exhausted and the point at which diseconomies of scale are encountered. Or, alternatively, economies and diseconomies of scale may be largely self-canceling over an extended range of output. Figure

[1] See Leonard W. Weiss, *Economics and American Industry* (New York: John Wiley & Sons, Inc., 1961), pp. 347–350.

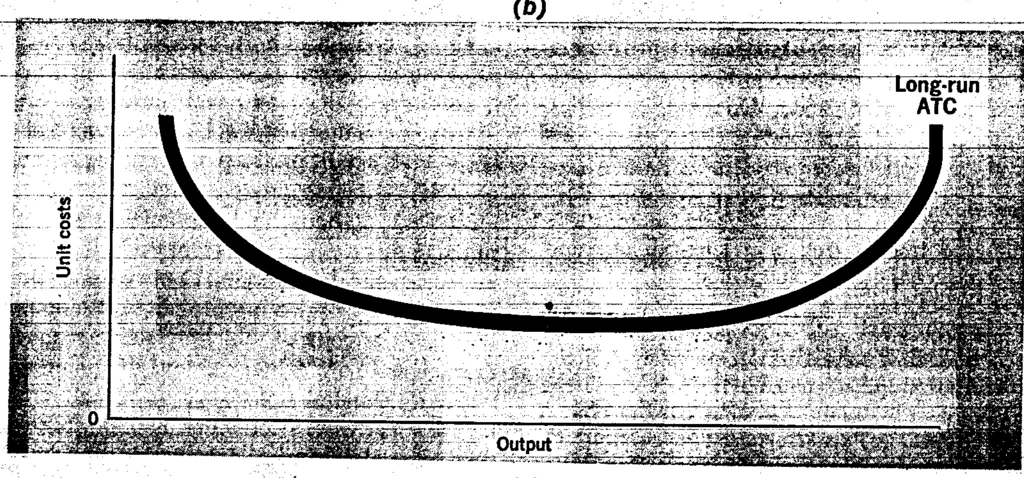
[2] Statement by George Stocking, cited in Walter Adams (ed.), *The Structure of American Industry*, 3d ed. (New York: The Macmillan Company, 1961), p. 180.


Unit costs
Long-run
ATC
0
Output

(a)


Unit costs
Long-run
ATC
0
Output

(b)


Unit costs
Long-run
ATC
0
Output

(c)

25–6*c* illustrates the situation in which average costs are relatively constant over a wide range of output.

Caution: We are not implying here that long-run unit costs are the only determinant of the structure of industry. Indeed, it was stressed in Chapter 22 that the major determinants of the competitiveness of industry are several and varied. We are saying that there is considerable evidence that cost considerations are one important force in determining the number and size of firms in a particular industry.

SUMMARY

1. Economic costs include all payments which must be received by resource owners in order to assure their continued supply in a particular line of production. This definition includes explicit costs which flow to resource suppliers who are separate from a given enterprise and also implicit costs which are the remuneration of self-owned and self-employed resources. One of the implicit cost payments is a normal profit to the entrepreneur for the functions he performs.

2. In the short run a firm's plant capacity is fixed. The firm can use its plant more or less intensively by adding or subtracting units of variable resources, but the firm does not have sufficient time to alter its plant size.

3. The law of diminishing returns describes what happens to output as a fixed plant is used more intensively. The law states that, as successive units of a variable resource such as labor are added to a fixed plant, beyond some point the resulting marginal product associated with each additional worker will decline.

4. Because some resources are variable and others fixed, costs can be classified as variable or fixed in the short run. Fixed costs are those which are independent of the level of output. Variable costs are those which vary with output. The total cost of any output is the sum of fixed and variable costs at that output.

5. Average fixed, average variable, and average total costs are simply fixed, variable, and total cost per unit of output. Average fixed costs decline continuously as output increases, because a fixed sum is being apportioned over a larger and larger number of units of production. Average variable costs are U-shaped, reflecting the law of diminishing returns. Average total cost is the sum of average fixed and average variable cost; it too is U-shaped.

6. Marginal cost is the extra, or additional, cost of producing one more unit of output. Graphically, marginal cost cuts ATC and AVC at their minimum points.

7. The long run is a period of time sufficiently long for a firm to vary the amounts of all resources used, including plant size. Hence, in the long run all costs are variable. The long-run ATC, or planning, curve is composed of segments of the short-run ATC curves, which represent the various plant sizes a firm is able to construct in the long run.

8. The long-run ATC curve is generally U-shaped. Economies of scale are first encountered as a small firm expands. A number of considerations—greater specialization in the use of labor and management, the ability to use the most efficient equipment, and the more complete utilization of by-products—contribute to these economies of scale. Dis-

FIGURE 25–6. VARIOUS POSSIBLE LONG-RUN AVERAGE-COST CURVES.
(*a*) When economies of scale are many and diseconomies remote, the ATC will fall over a wide range of production. (*b*) If economies of scale are few and diseconomies are quickly incurred, minimum unit costs will be encountered at a relatively low output. (*c*) Where economies of scale are rather rapidly exhausted and diseconomies not encountered until a considerably large scale of output has been achieved, long-run average costs will be relatively constant over a wide range of output.

economies of scale stem from the managerial complexities which accompany large-scale production.

9. The relative importance of economies and diseconomies of scale in an industry is often an important determinant of the structure of that industry. Generally speaking, where economies of scale extend to large levels of output, an industry tends to be comprised of a small number of large-scale producers. When economies of scale are exhausted at relatively low levels of output, there tends to be a large number of small firms in an industry.